* Introduction to fundamentals :

Standard data transformation for time series is logging the data then taking successive differences.

* Mean : $\mu_{x_t} = \mathbb{E}(x_t)$

$\mu_{w_t} = \mathbb{E}(w_t) = 0$ where $w_t \sim N(0,1)$ white noise

$\mu_{v_t} = \mathbb{E}(v_t) = \frac{1}{3}\left(\mathbb{E}(w_{t+1}) + \mathbb{E}(w_t) + \mathbb{E}(w_{t-1})\right) = 0$ where

v_t is the moving average series (smoothing series)

$\mu_{x_t} = \mathbb{E}(x_t) + \mathbb{E}\left(\delta_t + \sum_{j=1}^{t} w_j\right) = \delta_t + \sum_{j=1}^{t} \mathbb{E}(w_j)$

where x_t is a random walk with a drift

$\mu_{x_t} = \mathbb{E}\left(\cos\left(2\pi \frac{t+15}{50}\right) + w_t\right) = \cos\left(2\pi \frac{t+15}{50}\right)$ so

the mean of x_t is a cosine wave

* Autocovariance : $\gamma_x(s,t) = \mathrm{Cov}(x_s, x_t)$
$= \mathbb{E}\left((x_s - \mu_s)(x_t - \mu_t)\right)$

If $\gamma_x(s,t) = 0$ then x_t and x_s are not linearly dependent

$\gamma_w(s,t) = \begin{cases} \sigma^2, & s = t \\ 0 & s \neq t \end{cases}$ as $w_t \overset{iid}{\sim} N(0,\sigma^2)$

$$Cov\left(\sum_{j=1}^{m} a_j X_j, \sum_{k=1}^{r} b_k Y_k\right) = \sum_{j=1}^{m}\sum_{k=1}^{r} a_j b_k \, Cov(X_j, Y_k)$$

* Autocorrelation Function (ACF) : $\rho(s,t) = \dfrac{\gamma(t,s)}{\sqrt{\gamma(t,t)\,\gamma(s,s)}}$

* Stationary : 1) μ_t constant and independent of t
  2) $\gamma(s,t)$ depends on times t and s only

Random walk is not stationary as $\gamma(s,t) = \min(s,t)$
σ^2 depends on time.

When referring to a stationary time series process we can use $\mu = \mu_t$ as it is independent of t and x_t only depends on t and s so suppose $s = h + t$ where h is the time shift or lag.

$$\gamma(t, t+h) = Cov(x_t, x_{t+h}) = Cov(x_0, x_h) = \gamma(0, h)$$

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}, \quad \gamma(h) = Cov(x_{t+h}, x_t)$$

White noise and moving average are stationary

* Trend Stationarity : $x_t = \beta t + y_t$, y_t stationary

1) $\mu_{x_t} = E(x_t) = \beta t + \mu_y$

2) $\gamma_x(h) = Cov(x_{t+h}, x_t)$
$= \mathbb{E}((x_{t+h} - \mu_{x_{t+h}})(x_t - \mu_t))$
$= \mathbb{E}((y_{t+h} - \mu_y)(y_t - \mu_y))$
$= \gamma_y(h)$

3) $\gamma(0) = \mathbb{E}((x_t - \mu)^2) = Var(x_t)$

4) $\gamma(h) = \gamma(-h)$

* Wold Decomposition : Any stationary time series x_t can be written as linear comb. (filter) of white noise terms

$$x_t = \sum_{j\geq 0} \Psi_j w_{t-j} + \mu$$

$\sum_{j\geq 0} \Psi_j^2 < +\infty$ & $\Psi_0 = 1$, we call them linear processes.

* Jointly Stationary : If both x_t & y_t are stationary and their cross-covariance function is of lag h

$\gamma_{xy}(h) = Cov(x_{t+h}, y_t)$
$= \mathbb{E}((x_{t+h} - \mu_x)(y_t - \mu_y))$

* Ordinary Least Squares for time series :

* $x_t = \beta_0 + \sum_{j=1}^{q-1} \beta_j z_{tj}$

* $S = \sum_{i=1}^{n} w_i^2 = \sum_{i=1}^{n} \left(x_i - \beta_0 - \sum_{j=1}^{q-1} \beta_j z_{tj}\right)^2$

We solve for $\frac{\partial S}{\partial \beta_i} = 0 \;\; \forall i=0,\dots,q-1$ to get the coef.

* $SSE = \sum_{i=1}^{n} (x_i - \hat{x}_i)^2$

* $\mathbb{E}(MS_{Res}) = \mathbb{E}\left(\frac{SSE}{n-q}\right) = \sigma^2$

* $t = \frac{\hat{\beta}_j - \beta_j}{SE(\beta_j)}$, in OLS $SE(\beta_j) = \sqrt{(X^T X)^{-1}_{jj} \, MS_{Res}}$

* $F = \frac{MS_{Reg}}{MS_{Res}}$

* Coefficient of determination: $R^2 = \frac{SS_{Reg}}{SS_T}$

* Akaike's Information Criterion: $AIC = \log\left(\frac{SSE_K}{n}\right) + \frac{n+2K}{n}$

* Bayesian Information Criterion: $BIC = \log\left(\frac{SSE_K}{n}\right) + \frac{K \log n}{n}$

* AIC is better with small samples while BIC is superior in larger samples.

* Suppose the data is not stationary:

1) Detrending: $x_t = \mu_t + y_t$, $\mu_t = \beta_0 + \beta_1 t$

$$\hat{y}_t = \hat{x}_t - \hat{\mu}_t$$

2) Differencing: $\nabla x_t = x_t - x_{t-1}$ removes linear trend

$\nabla^2 x_t = x_t - 2x_{t-1} + x_{t-2}$ removes quadratic trend

Differencing of order d : $\nabla^d = (1-B)^d$

$$\nabla^3 x_t = (1-B)^3 x_t$$
$$= (1 - 3B + 3B^2 - B^3)\, x_t$$
$$= x_t - 3B(x_t) + 3B^2(x_t) - B^3(x_t)$$
$$= x_t - 3x_{t-1} + 3x_{t-2} - x_{t-3}$$

3) Heteroscedasticity : $y_t = \log(x_t)$ transformation

4) Box - Cox : $y_t = \begin{cases} \dfrac{x_t^{\lambda} - 1}{\lambda} , & \lambda \neq 0 \\ \log(x_t) , & \lambda = 0 \end{cases}$

* Classical Structural Modeling :

We decompose $x_t = \underset{\text{trend}}{T_t} + \underset{\text{seasonal}}{S_t} + \underset{\text{noise}}{N_t} + \underset{\text{cyclic}}{C_t}$

* ARIMA Modelling

* Autoregressive model AR(p) :

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t$$
where x_t is stationary and w_t is white noise

* The models $x_t = \pm x_{t-1} + w_t$ are not AR as they are a random walk which is not stationary

If x_t is an AR model then $|\phi| < 1$ $\quad x_t = \sum_{j=0} \phi_j w_{t-j}$

* Causal time series : $x_t = \mu + \sum_{j \geq 0} \phi_j w_{t-j}$

* Moving Average Model of order q :

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q}$$ always stationary

* Invertible time series : $w_t = \sum_{j \geq 0} \pi_j x_{t-j}$; $\sum_{j \geq 0} \pi_j^2 < +\infty$

* ARMA(p,q) : $x_t = \alpha + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \dots + \theta_q x_{t-q}$
where x_t is causal and invertible

$$x_t = \beta_0 + \beta_1 x_{t-1} + \dots + \beta_p x_{t-p} + \epsilon_t$$ where
$$\epsilon_t = w_t + \theta_1 w_{t-1} + \dots + \theta_q x_{t-q}$$

ϕ is restricted to obtain causality
θ is restricted to obtain inventibility

* Model redundancy :

$x_t = 0.3 x_{t-1} + 0.4 x_{t-2} + w_t + 0.5 w_{t-1}$ looks like ARMA(2,1)
$(x_t - 0.3 x_{t-1} - 0.4 x_{t-2}) = w_t + 0.5 w_{t-1}$
$(1 - 0.3B - 0.4B^2) x_t = (1 + 0.5B) w_t$
$(1 + 0.5B)(1 - 0.8B) x_t = (1 + 0.5B) w_t$
$(1 - 0.8B) x_t = w_t$

Hence, $x_t = 0.8 x_{t-1} + w_t$ which is AR(1)

We can check this on R by observing the roots of the polynomial.

* Partial Autocorrelation Function (PACF) :

$$\phi_{hh} = \rho(x_h - \hat{x}_h, x_0 - \hat{x}_0) \quad , \quad h \geq 2$$

$$\phi_{11} = \rho(x_1, x_0)$$

where $\hat{x}_h$ is the regression of x_h on $\{x_1, \ldots, x_{h-1}\}$
and $\hat{x}_0$ is the regression of x_0 on $\{x_1, \ldots, x_{h-1}\}$